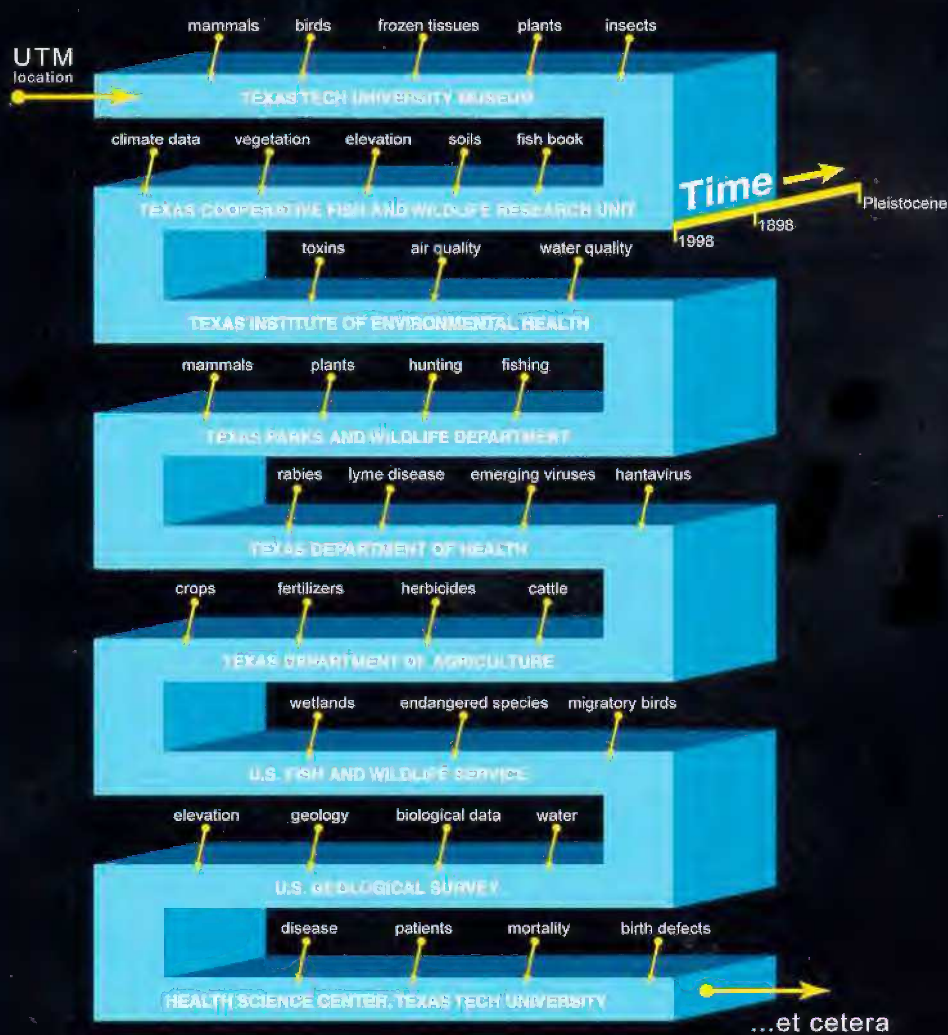
# Occasional Papers

Museum of Texas Tech University

Number 186 21 December 1998


# Bioinformatics:

## A Multidisciplinary Approach for the Life Sciences

## Executive Summary

- Bioinformatics provides a framework for understanding biocomplexity.
- Texas Tech University has been in the process of developing a bioinformatics program since the early 1990s when the term bioinformatics was first used.
- The Natural Science Research Laboratory of the Museum of Texas Tech University is among the first to develop computer-based catalogues and to use bar-coded tags for recording information on specimens.
- Texas Tech University, in concert with Texas Parks and Wildlife Department, initiated a computerized Natural Science Database in 1994 as an extension of the computerized collections databases developed by the Museum in the early 1980s.
- The emerging bioinformatics program of Texas Tech University involves the multidisciplinary work of at least 58 individuals in 10 departments and 4 colleges.
- Relational databases being developed are compatible with the National Biological Information Infrastructure of the U.S. Geological Survey, Biological Resources Division.
- Databases and research products are being placed on the World Wide Web at <www.tcru.ttu.edu/tcru/> and linked to the Museum web site at <www.nsrl.ttu.edu>.
- Relational databases are being developed to provide organization and structure for disparate data such things as mammals, habitats, vegetation, soils, precipitation, elevation, and photo-documentation.
- Society will benefit from data being arranged for visual presentation in geographic information systems, maps, reports, tables, and graphs.
- Historic data and archived specimens provide the base against which we measure change and identify voids in scientific knowledge.
- Well-structured relational databases will serve economic development by providing rapid information on which sound decisions may be based.
- Students trained in developing relational databases readily find employment in the information industry.

**Front cover:** Schematic representation of a relational database consisting of multiple databases housed in a number of departments at Texas Tech University and agencies of the State of Texas. Data from any one data set— rabies, for example— can be linked in time and space (Universal Transverse Mercator, UTM) to reveal the relationships of mammals which carry rabies to habitat (vegetation, land use, crops, pesticides, climate, geology, etc.) and impacts on society (human health and economics).

# Bioinformatics: A Multidisciplinary Approach for the Life Sciences

*Nick C. Parker, Robert D. Bradley, John M. Burns, Gary F. Edson, Donald R. Haragan, Clyde Jones, R. Richard Monk, John T. Montford, Carleton J. Phillips, David J. Schmidly,* and *Robert J. Baker*

Bioinformatics is a new field of science that utilizes data from the biological sciences and many ancillary disciplines (Fishman, 1996; Grace, 1997). A formal definition of bioinformatics has been given as " the systematic development and application of computing systems and computational solution techniques analyzing data obtained by experiments, modeling, database search, and instrumentation regarding biological aspects" (<http://www.bioplanet.com/whatis.html>). We have broadened and extended this definition to include "the delivery of the data and its synthesis to potential users." By this definition, we envision the delivery of data and products developed from a synthesis of data to school children, educators, landowners, personnel in state and federal agencies, legislators, lawmakers, investors, bankers, and other members of the public.

Although in its infancy, bioinformatics is a rapidly growing field. In a search of the Internet, seven search engines (Excite, Lycos, Infoseek, Yahoo, HotBot, Alta Vista, and Northern Lights) identified a 1993 document from a 1992 symposium as the first to contain the word bioinformatics, defined as the management of biological information through computer technology (Lim et al., 1993). By 1997, there were 843 documents with the term bioinformatics identified by these same seven search engines. As an indication of the rapid growth of interest, the search engine Northern Lights located 23,685 documents containing the term bioinformatics in November 1998. The use of this term has primarily been associated with molecular genomics. The Human Genome Project, with its goal of determining the sequence of approximately three billion base pairs, made it imperative that computer technology be utilized to organize and manage the explosion of genomic information (Collins et al., 1998). Today, it is possible to search the database of DNA sequences of nearly one billion nucleotide bases to determine the recorded sequences that most closely match any newly sequenced DNA fragment. Such a search typically takes a few minutes and frequently only a few seconds.

In our world of specialists and special interest groups, it's easy to become immersed in the details and lose sight of the big picture. We tend to concentrate on problems in our own back yard and leave problems in other parts of the world to others. Biologist E.O. Wilson in his recent book "Consilience: The Unity of Knowledge" presents a strong argument that all knowledge is related and that the world operates in accordance with a number of fundamental natural laws (Wilson, 1998*a*, 1998*b*). The relatedness of biological life, the intricacies of the environment, and the importance of biodiversity has been recognized by others. Rita Colwell, Director of the National Science Foundation, promotes use of the term biocomplexity to denote the variety of the living planet and our ability to understand its intricately integrated systems of life (Mervis, 1998). Decisions made by society are not made in isolation of the natural world, but often are shaped by the biological world, the flora and fauna, in turn reflecting the chemical and physical condition of the environment. Thus, bioinformatics provides a framework for understanding this biocomplexity.

Bioinformatics is relevant to all of us, as we all live in the biological world. The data sets upon which bioinformatics are based help us understand, recognize, and address such crucial issues as environmental quality, biodiversity, habitat composition, resource allocation, and sustainable development opportunities. The organization of biological data and ancillary ecological attributesinto relational databases provides an order of understanding not available prior to the development of bioinformatics. It is our view that the accessibility and interpretation of these data will have an unparalleled effect on the utility of biological data not only in the sciences, but in the private sector as well.

Insight from such data provides knowledge about the world's biocomplexity, biodiversity, and maintenance of aesthetics, as well as the ability for sustainable plan-

Parker et al., 1998. Bioinformatics: A Multidisciplinary Approach for the Life Sciences. Occasional Papers, Museum of Texas Tech University 186:i+1-8

ning and economic opportunity. In a world with 1.75 million species of organisms described and estimates ranging between 4 and 40 + million undescribed species, extinction has reached an alarming estimated rate of 30,000 to 300,000 species per year (Wilson, 1988). These are species with intrinsic value as the "cogs in the wheel" (the simple inter-relatedness of organisms in the ecosystem) and of unknown potential value as a source of medicinal drugs, chemicals, DNA sequences, and alternate food crops.

## Existing Programs in Bioinformatics

Few bioinformatics programs now exist in the United States and those that do are devoted almost exclusively to genomic projects. The National Center for Genome Resources contains the Genome Sequence database, established to increase awareness of bioinformatics by medical specialists and sub-specialists. The U.S. Department of Agriculture has established the Agricultural Genomics Program employing bioinformatic techniques. The U.S. Department of Energy has a bioinformatics division deeply involved with the Human Genome Project. Another example in the Federal Government is in the Department of Interior which has established the National Biological Information Infrastructure (NBII), a bioinformatics program of another name managed by the U.S. Geological Survey. The Presidential Committee of Advisors on Science and Technology recommended that the Federal Government invest a minimum of $40 million per year for the next 5 years to develop the next generation NBII. At the state level, the Southwest Region of the U.S. Fish and Wildlife Service, in cooperation with the Center for Wildlife Law at the University of New Mexico Institute of Public Law, has developed an Internet linkage titled 'the Southwest Roadmap' (<http://roadmap.unm.edu>) to link multiple websites of interest to natural resource managers and researchers. Several pharmaceutical companies in the United States have established departments or divisions with bioinformatics as part of the formal name. For example, the Immunex Corporation, Seattle, Washington, has a Department of Bioinformatics; Hoffmann-La Roche, Nutley, New Jersey, has the Department of Oncology and Bioinformatics, and SmithKline Beechem Pharmaceuticals, Collegeville, Pennsylvania, has a division of Bioinformatics.

The biomedical community of Europe and much of the world seems to have embraced and implemented programs employing bioinformatics. For example, the Max-Delbrück Center, Berlin, Germany, has established the Department of Bioinformatics, the Swiss Institute for Environmental Cancer Research, Lausanne, Switzerland, has a Bioinformatics Group; the National University of Singapore has established a Bioinformatics Center in the National University Hospital; and South Africa has the National Bioinformatics Institute. In Mississauga, Ontario, Canada, the Base 4 Bioinformatics Company has been established to increase understanding in (1) the basic biological processes of the body, (2) ways in which these processes may malfunction to cause disease, and (3) improved drug discovery and development processes.

Although it is clear that bioinformatics is a growing field and that several state, federal, and international agencies are beginning to implement programs in this area, what is needed is an academic training program that prepares and trains scientists for these programs (Anonymous, 1996; Marshall, 1996*a*; Marshall, 1996*b*). For example, any recent issue of *Science* contains job vacancy announcements for six to twelve positions in bioinformatics. In Science 281 (5385), the University of Missouri, Clemson University, and the University of Georgia advertized for bioinformatics positions to work in maize genomics. In the same issue, the National Cancer Institute advertized for a database administrator in the Office of Informatics and the University of Denmark advertized five positions in bioinformatics. Qualifications required for these jobs typically include backgrounds in biology and computer sciences, knowledge of Oracle and other database systems, skills in computer techniques, telecommunications, mapping techniques, development of world wide web pages, and the ability to communicate effectively with diverse users of the data.

## Current Activities at Texas Tech University

The basis of a bioinformatics program at Texas Tech University has been developing for a number of years. A 1996 publication with 23 authors from 13 different sectors (departments, universities, state, and federal agencies) outlined the critical need for a natural science database and its value to resource management and public health (Baker et al., 1996). In September of 1998, a presentation on bioinformatics was included as part of the program of the Taxonomy Database Working Group in Reading, England (Parker et al., 1998). These two publications represented the combined efforts of dozens of students, staff, and faculty. For example, in addition to the multidisciplinary author lines, these publications represent the compiled work of at least 58 individuals in 10 departments and 4 colleges at Texas Tech University (see acknowledgments), all of whom were necessary to solve the questions at hand.

At Texas Tech University, we planned to create a relational database which would extend across the University to many colleges and departments. The Museum, through the Natural Science Research Laboratory, in partnership with the Texas Cooperative Fish and Wildlife Research Unit (a component of U.S. Geological Survey's Biological Resources Division) and Texas Parks and Wildlife Department, have taken the lead in initiating this program. This partnership program has state (Texas GAP; Parker et al., in press), national (National GAP Analysis Program; Scott et al., 1993), and international (Rio Grande GAP, headquartered at TTU; Gonzalez-Rebeles et al., 1997) components. To serve the components, we have acquired the database Oracle to provide the data structure for several databases currently residing in the Museum and in the GAP lab. Data to be placed in the Oracle database and selectively made available through the Web include vegetation maps for each county in Texas, maps of vertebrate distributions in Texas, records from the mammal collection, text, photos, and maps from <u>The Mammals of Texas</u> (Davis and Schmidly, 1994), and climate data from 3,860 sites in Texas, some with daily records spanning over 100 years.

A doctoral student currently working on the historical distribution and population size of scaled quail is using vegetation maps of Texas-GAP and other ancillary data to quantify changes in habitat with the population size of scaled quail. The two data sets included in this study are climate and the Breeding Bird Survey. The climate data are from 3,860 sites in Texas with daily records of five to ten variables including average, maximum, and minimum temperature, rainfall, and snow. Some of these records span over 100 years. These data, purchased by Civil Engineering and used for their research, were in multiple files on a CD. Each data point was stored as a separate file on the CD. For 12 months, students in the TX-GAP lab opened these files and transferred data point-by-point into a relational database allowing analysis of these datain a geographical information system (GIS). These data are now in a format that can be used in the GAP lab and by others for spatial analysis.

The Breeding Bird Survey is maintained by the Biological Resources Division of the U.S. Geological Survey (formerly by the U.S. Fish and Wildlife Service). The data for Texas are available on five microfilm tapes and cover the years 1967 through 1993. The 1994 through 1996 data are available in hardcopy format. The 1997 data are being computerized by the national office of the Breeding Bird Survey. The time estimate for transferring the 1967 through 1993 data for Texas into computer files is 18 months at 40 hours per week. The student needing these data will be able to computerize only a small sample of the Breeding Bird Survey for Texas. However, once these selected data are in a relational database, we expect to add other sites, and years, as components of new projects. We also are negotiating with the Breeding Bird Survey office to secure funding to assist in computerizing these files.

The aforementioned database is being developed for broad-scale, as well as specific application use by Texas Parks and Wildlife Department as the scientific basis for a natural resource management plan. The plan is to collect and properly archive samples of the vertebrate forms so that a detailed and accurate record will be available for future questions concerning the distribution and abundance of species at the beginning of the 21st century. The geographic localities being sampled are the properties administered by TPWD. Although these properties are not randomly distributed across the state, there is representation of most ecological regions habitat types. In the archives of the NSRL, there are samples of properly prepared and documented classical museum specimens as well cryopreserved samples of liver, heart, kidney, muscle, spleen, and lungs. At this point we know that each specimen contains vast amounts of information that can be valuable to science and society related to topics such as genetics (especially DNA sequences), conservation genetics, disease, levels of contamination of tissues from pollution, and presence of pathogens and viruses. We realize that this biological information is critical for implementing wise decisions; however, what we don't know is how much new technology will allow us to study relationships which have not been conceived. If the technological advances of the next century or the next 25 years match the last century or even the last quarter century, then the extractable knowledge will experience an exponential growth rate (Corn and Horrigan, 1984).

## The Vision

Our goal is to develop a relational database at Texas Tech University with spatial data including information about ecoregions (Bailey et al., 1994), climate, soils, distribution of vertebrates, vegetation, crops, pesticide application, field notes, photos, maps, land-use, land-cover, parks, public properties, and wildlife management areas. Electronic databases also are planned to include 3-D digital images of essentially all objects in museum collections, including items such as skulls and arrowheads and using tools now available in CAD-CAM programs, these images could be rotated and viewed from any angle. Features undetectable with the unaided eye could be computer-enhanced to allow analysis and comparisons that could not previously be made, even when holding the real object in hand. Similarly, in applications extending beyond the Museum 3-D molecular models of DNA, proteins, and other biochemicals would aid in development of drugs and understanding of biological processes. Interactions of antigens and antibodies, hormones and receptor sites, and actions of endocrine disruptors would be enhanced by 3-D images that could be viewed from all angles as holograms are viewed.

The 3-D technology now available in the entertainment field will soon be available as a tool for science. Native American pictographs or rock drawings are now stored as 3-D images at the University of Arkansas. Other museums also may be using similar technology for a variety of collections. Improvements in computer hardware, software, and Internet transmissions (Miller, 1998; Seifa 1996; Williams, 1997) will make this an attractive method for presenting information on museum collections to viewers around the world. The University of Illinois at Chicago has developed a Cave Automated Virtual Environment (CAVE); a virtual reality environment in which the viewer can walk around and interact with visual images. Home entertainment systems of the future will likely include surround video along with surround sound. This virtual image technology could also be used in the medical sciences to view molecular structure, drugs and receptor sites, MRI images, and provide visual evaluation of prostheses to be placed in the body or to view organs prior to surgery. IBM's new Pacific Blue supercomputer, today's fastest, with speeds of 3.9 trillion calculations per second, will allow image analysis and processing never before imagined -- and most likely the world's fastest computer today undoubtedly will be far too slow for tomorrow's needs. With state-of-the-art technology, we envision development of CDs and other storage media to contain selected datasets that will serve schools, businesses, and the public at sites not connected to the Internet, or at sites where hardware limits the speed of Internet accessability. We envision the development and delivery of products to be an on-going process to keep abreast of technological advances. The storage media of today will probably soon become obsolete as society demands faster and more user-friendly data synthesis and delivery.

## The Value

A robust relational database would provide structure for data now available and expanding at an unprecedented rate. For example, the gross state product for Texas was $372 billion in 1990, and in 1995 the construction industry was valued at $91 billion (Ramos and Plocheck, 1995). In 1994, agriculture products at the farm and ranch gate level were valued at $13 billion, and at $42 billion for the collective agricultural industries. All of these industries are regulated at the state and federal level in some manner to control, for example, air pollution, soil erosion, effluents, and site selection for specific projects. Data required for farmers, developers, and other businesses to complete permit applications include site surveys for threatened and endangered species, evaluation of wetlands status, potential for soil erosion, location of aquifers, groundwater recharge zones, and location of existing wells for water, oil, and gas. The location of hazardous waste sites, landfills, and the storage sites for toxic material must be known before many farming, development or even conservation projects can be permitted.

A relational database that brings together many data sets would be of great value to society. For example, a database providing the information needed to complete the applications for the permits required in the construction industry alone could save the industry $1 million per day in interest if only one tenth of one percent of the industry could reduce permit application time by 1 day [$10 billion x 0.01 (fraction of industry benefitting) x 0.10 (interest rate) = $1 million]. Similar savings in the agricultural industry would save over $4 million per day. No dollar amount has been assigned to the potential value of such a database containing information about distribution of rabies, hantavirus, and other emerging viruses. The expected value of such a database to protect environmental quality and public health would be immense.

## Summary

The next millennium has been labeled the information age and scientific data must be readily available as the basis for decisions affecting our growing society. Advancements in communications, computers, remote sensing, and instantaneous analysis are rapidly pushing science to new fields and forcing a merger of fields across interdisciplinary lines. For example, as human health is placed at risk by pollution and serious diseases such as HIV, Ebola, and other emerging viruses, the solutions are not to be found in one discipline of science, but in a synthesis from multiple disciplines.

We have established a bioinformatics program at Texas Tech University that involves the Museum and other multidisciplinary fields and programs, such as the Texas GAP program and the Department of Biological Sciences. One goal of this program is to build a relational database, as a linkage of distributed databases, which can be accessed through the Internet. We are using Oracle as the primary relational database with specific existing applications running in Visual FoxPro, MS Access, and Excel. Examples of the data now available include vertebrate collections from the Museum; field notes, photos and records of the 1895-1906 biological survey of Texas; routes of the Breeding Bird Survey of Texas conducted annually by the U.S. Fish and Wildlife Service; daily weather data from 3,860 sites in Texas with some records extending back over 100 years; current landscape photographs of habitats in Texas; soil maps for Texas; vegetation in Texas; aerial videography; digital elevation maps; and Landsat TM scenes for all of Texas and the North American Trade Zone, along the border with Mexico, established by the North American Free Trade Agreement.

Several products have been produced to date and placed on the Internet. Applications include soil maps by agriculture crop type, design and placement of constructed wetlands for extraction of nutrients from the effluent of cattle feedlots, integration of aquaculture with traditional agriculture, distribution models for vertebrate species in Texas, and identification of areas with high levels of biodiversity. These and other projects will allow researchers, resource managers, landowners, school children, and the general public to use the best available data in their livelihoods, projects, research, and decisions.

## Acknowledgments

A project of this scope is not conducted without the direct input of a large number of authors and the underlying support of an even larger number of students, staff, faculty, and colleagues. This work was sponsored by Texas Tech University in cooperation with Texas Parks and Wildlife Department to whom we give special thanks to the Director, Andrew Sansom. Each individual listed played a crucial role in development of the program reported in this paper. We thank them all and apologize in advance for those who should have been listed but were overlooked. Any omissions were solely the responsibility of the senior author. The Texas Cooperative Fish and Wildlife Research Unit is jointly sponsored by Texas Parks and Wildlife Department, Texas Tech University, U.S. Geological Survey-Biological Resources Division, and the Wildlife Management Institute.

We thank by college and department the following:

**College of Arts and Sciences**, Museum of Texas Tech University— Brian D. Bisbee, Lisa Bradley, Jessica M. Carrillo, Susan Fishman-Armstrong, Burhan M. Gharaibeh, Britney J. Hager, Kristina Halcomb, Amy S. Halter, Raegan King, Shannon Lenerose, Christena Parker-Stephens, Frank D. Yancey; Department of Biological Sciences— Kelly Allen, Darin Bell, Darin Carroll, Larry L. Choate, Jim R. Goetze, Melinda Clary, Seth Davis, Cody Edwards, Andrea Ernst, Sara Hrachovy, Chad Hutcherson, Ted Jolley, Mark O'Neill, Nicole Lewis-Oritt, Stacy Mantooth, Roslyn Martinez, Cole Matson, Charlene Fletcher, Lottie Peppers, Brenda Rodgers, Scott Schrader, Irene Tiemann-Boege, Jeffrey Wickliffe, Raigan Wilson, Zhaoxia Zhang; Interdisciplinary Studies— Oleksiy Knyazhnitskiy, Qing Xia.

**College of Agricultural Sciences and Natural Resources**, Department of Range, Wildlife, and Fisheries Management— T.J. Cox, Eric Holt, Raquel Leyva, Wade Shoop; Texas Cooperative Fish and Wildlife Research Unit— Joseph Balderas, Carlos Gonzalez-Rebeles, Sheri Haskell, Nancy Hubbard, Guoding Kang, Min Li, Maricruz Longoria, Jodi Martin, Reynaldo Patiño, and Leah Simon; Landscape Architecture— Jeanette Pritchard, Eric Schultz.

**College of Engineering**, Civil Engineering— Clifford B. Fedler, John Domansky, Eli Lerma, Troy White; Engineering Technology— Kevin Egan, Earl Farley, Kevin Fisher; Computer Science— Hui Chi, Hye Kyoung Kang, Joe Salinas, Travis Teague.

**College of Business Administration**, Information Systems and Quantitative Sciences— John Durrette, Bin Mai, Nirup Menon, Surya Yadav.

Supported by National Science foundation grant number DBI-9808928, R. D. Bradley, PI.

## Literature Cited

Anonymous, 1996. Bioinformatics: New frontier calls young scientists. Science 273(5272):265.

Bailey, R.G., P.E. Avers, T. King, and W.H. McNab (eds.), 1994. Ecoregions and subregions of the United States. Colored map (1:7,500,000) with supplementary table of map unit descriptions, compiled and edited by W.H. McNab and R.G. Bailey. USDA Forest Service, Washington, DC.

Baker, R.J., R.A. Albin, R.D. Bradley, J.J. Bull, J. Burns, K.A. Clark, G. Edson, R.E. Estrada, E. Farley, C.B. Fedler, B.M. Gharaibeh, R.L. Hammer, C. Jones, R.R. Monk, J.T. Montford, G. Moore, N.C. Parker, J. Rawlings, A. Sansom, D.J. Schmidly, R.W. Sims, H. Wichman, and F.D. Yancey, 1997. Natural Science Database: Resource management and public health. Pages 10-20 *in* M. Shaughnessy (ed.) Collaboration: The 'Key' to Success. Proceedings of the 4th Annual Conference, Organization of Fish and Wildlife Information Managers. Key Largo, FL. 96 pp.

Collins, F.S., A. Patrinos, E. Jordan, A. Chakravanti, R. Gesteland, L. Walters, and members of the DOE and NIH planning groups, 1998. New goals for the U.S. Human Genome Project: 1998-2003. Science 282:682-689.

Corn, J.J. and B. Horrigan, 1984. Yesterday's tomorrows. Summit Books, New York.

Davis, W.B. and D.J. Schmidly, 1994. The Mammals of Texas. Texas Parks and Wildlife Department. Austin, Texas. 338 pp.

Fischman, J. 1996. Bioinformatics: Working the Web with a virtual lab and some Java. Science 273:591

Gonzales-Rebeles, C., N.C. Parker, V. Burke, M. Jennings, and G. Ceballos, 1997. First international Gap Analysis project. Gap Analysis Bulletin No. 6. USGS/BRD Gap Analysis Program, Moscow, Idaho.

Grace, J.B. 1997. Bioinformatics: Mathematical challenges and ecology. Science 275:1861.

Lim, H.A., J.W. Fickett, C.R. Cantor, and R.J. Robbins (eds.), 1993. Bioinformatics, supercomputing and complex genome analysis. World Scientific Publishing Co., New Jersey. 648 pp.

Marshall, E. 1996*a*. Bioinformatics: Hot Property: Biologists who Compute. Science 272:1730.

Marshall, E. 1996*b*. Bioinformatics: Demand outstrips supply. Science 272:1731.

Mervis, J. 1998. The biocomplex world of Rita Colwell. Science 281:1944-1947.

Miller, C. 1998. Penny-sized CD. Popular Science, May 1998:35.

Parker, N.C., R.D. Bradley, G. Edson, C. Jones, R.R. Monk, D.J. Schmidly, and R.J. Baker, 1998. Bioinformatics: An integrated program for academia, resource managers, and the public. <http://www.tcru.ttu.edu/bioinfindex.html>

Parker, N.C., R.J. Baker, R.D. Bradley, C. Jones, R.R. Monk, D.J. Schmidly, R.W. Sims, and F.D. Yancey, II, in press.

Texas Tech Museum and Texas Gap Program: A partnership providing field data for GAP analysis. Gap Analysis bulletin No. 7. USGS-BRD Gap Analysis Program, Moscow, Idaho.

Ramos, M.G. and R. Plocheck (eds), 1995. 1196-1997 Texas Almanac. The Dallas Morning News, Inc., Dallas, Texas. 672 pp.

Scott, J.M., F. Davis, C. Csuti, R. Noss, B. Butterfield, S. Caileo, C. Groves, T.C. Edwards, Jr., J. Ulliman, H. Anderson, D., F.D. Erchia, and R.G. Wright, 1993. Gap Analysis: A geographic approach to protection of biological diversity. Wildlife Monographs 123:1-41.

Seifa, C. 1996. Bioinformatics: Do Java users live dangerously? Science 273:592.

Williams, N. 1997. Bioinformatics: How to get databases talking the same language. Science 275:301.

Wilson, E.O. 1988. Biodiversity. National Academy Press, Washington, DC. 521 pp.

Wilson, E.O. 1998*a*. Consilience: The unity of knowledge. Knoph, New York. 332 pp.

Wilson, E.O. 1998*b*. Integrated science and the coming century of the environment. Science 297:2028-2049.

*Addresses of authors:*

NICK C. PARKER

*Texas Cooperative Fish and Wildlife Research Unit*
*U. S. Geological Survey-Biological Resources Division,*
*Texas Tech University*
*Lubbock, TX 79409-2120*
*e-mail: nparker@ttu.edu*

ROBERT D. BRADLEY AND CLYDE JONES

*Department of Biological Sciences and Natural Science Research Laboratory, Museum of Texas Tech University,*
*Lubbock, TX 79409-3191*
*e-mail: izrdb@ttacs.ttu.edu (RDB),*
*cjones@packrat.musm.ttu.edu (CJ)*

JOHN M. BURNS

*Provost, Texas Tech University*
*Lubbock, TX 79409-2019*
*e-mail: john.burns@ttu.edu*

GARY F. EDSON

*Director, Museum of Texas Tech University*
*Lubbock, TX 79409-3191*
*e-mail: mxgfe@ttacs.ttu.edu*

DONALD R. HARAGAN

*President, Texas Tech University*
*Lubbock, TX 79409-2005*
*e-mail: don.haragan@ttu.edu*

R. RICHARD MONK

*Natural Science Research Laboratory*
*Museum of Texas Tech University*
*Lubbock, TX 79409-3191*
*e-mail: rmonk@ttu.edu*

JOHN T. MONTFORD

*Chancellor, Texas Tech University*
*Lubbock, TX 79409-2013*
*e-mail: jt.montford@ttu.edu*

CARLETON J. PHILLIPS

*Chair, Department of Biological Sciences*
*Texas Tech University*
*Lubbock, TX 79409-3131*
*e-mail: bhcjp@ttacs.ttu.edu*

DAVID J. SCHMIDLY

*Natural Science Research Laboratory*
*Museum of Texas Tech University*
*Lubbock, TX 79409-3191*
*e-mail: schmidly@ttu.edu*

ROBERT J. BAKER

*Department of Biological Sciences and Natural Science Research Laboratory, Museum of Texas Tech University,*
*Lubbock, TX 79409-3191*
*e-mail: rjbaker@ttu.edu*

## PUBLICATIONS OF THE MUSEUM OF TEXAS TECH UNIVERSITY

It was through the efforts of Horn Professor J Knox Jones, as director of Academic Publications, that Texas Tech University initiated several publications series including the Occasional Papers of the Museum. This and future editions in the series are a memorial to his dedication to excellence in academic publications. Professor Jones enjoyed editing scientific publications and served the scientific community as an editor for the Journal of Mammalogy, Evolution, The Texas Journal of Science, Occasional Papers of the Museum, and Special Publications of the Museum. It is with special fondness that we remember Dr. J Knox Jones.

Institutional subscriptions are available through the Museum of Texas Tech University, attn: NSRL Publications Secretary, Box 43191, Lubbock, TX 79409-3191. Individuals may also purchase separate numbers of the Occasional Papers directly from the Museum of Texas Tech University.

Layout and Design: R. Richard Monk
Cover Design: Eric Schultz and Jeanette Pritchard



This book was set in Times New Roman and printed on acid-free paper that meets the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources.

Printed: 21 December 1998

---

Library of Congress Cataloging-in-Publication Data

Occasional Papers, Number 186
Series Editor: Robert J. Baker

BIOINFORMATICS: A MULTIDISCIPLINARY APPROACH FOR THE LIFE SCIENCES

By: Nick C. Parker, Robert D. Bradley, John M. Burns, Gary F. Edson, Donald R. Haragan, Clyde Jones, R. Richard Monk, John T. Montford, Carleton J. Phillips, David J. Schmidly, and Robert J. Baker

ISSN 0149-175X

Museum of Texas Tech University
Lubbock, TX 79409-3191 USA
(806)742-2442

SEAL OF TEXAS TECH UNIVERSITY ™